Health and mortality of the elderly: the grade of membership method, classification and determination

Portrait, F.R.M.; Lindeboom, M.; Deeg, D.J.H.

1999

document version
Early version, also known as pre-print

.it der Economische Wetenschappen en Econometrie

Serie research memoranda

Health and Mortality of the Elderly: the Grade of Membership method, Classification and Determination

France Portrait
Maarten Lindeboom
Dorly Deeg

Research Memorandum 1999-22 **April 1999**

/a/l/e/r/t/

applied
labour
economics
research
team

vrije Universiteit amsterdam

Health and Mortality of the Elderly : the Grade of Membership method, Classification and Determination

France Portrait?
Maarten Lindeboom[2]
Dorly Deeg[3]
4

April 6, 1999

'Department of Econometrics, Free University
[2]Department of Economics, Free University and Tinbergen Institute
[3]Longitudinal Aging Study Amsterdam, Faculty of Medicine and Faculty of Social and Cultural Sciences, Free University
[4]The Longitudinal Aging Study Amsterdam (LASA) kindly allowed us to use their data. The LASA study is mainly financed by a long term grant of the Dutch ministry of Health, Welfare and Sports.
This paper has benefitted substantially from the comments and suggestions of Nol Merkies, Martin Foster, Owen O'Donnell and two anonymous referees of this journal. We furthermore acknowledge the valuable comments and suggestions of the participants of the seventh European workshop on Econometrics and Health Economics.
Address for correspondence : Department of Econometrics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
Email : FPortrait@econ.vu.nl

Abstract

With the aging of the society, issues concerning the reform of the Dutch health care system are ranked high on the political agenda. Sensible reforms of the health care system for the elderly require a thorough understanding of the health status of the old and of its dynamics preceding death. The health status of the elderly is intrinsically a multidimensional and dynamic concept and a rich set of indicators is needed to capture this concept in its full extent. This feature of health requires techniques to reduce dimensionality as it will in general be difficult to handle simultaneously all indicators in any economic analysis. In the first part of this paper we focus on methods that comprise these multidimensional measures into a limited number of indices. The Grade of Membership approach introduced by Manton and Woodbury in 1982 is specifically designed to characterize the complex concept of health. The method simultaneously identifies all dimensions of the concept of interest and the degrees to which an individual belongs to each of these types. We apply the method on a set of 21 indicators from a rich database of the Longitudinal Aging Study Amsterdam. The individual degrees of involvement in the different health dimensions obtained from this method are used in subsequent analyses of health and mortality.

1 Introduction

There have been significant increases in life expectancy of the population for most OECD countries and the trend towards aging of the society will become more prominent when the post World War II baby boomers reach the more advanced ages. In the Netherlands, in the mid nineteen fifties, about 7% of the population consisted of people aged 65 and above. Now four decades later this fraction has doubled and it is expected to become close to 25% in 2035. This growth of the elderly population in itself is expected to have a large impact on many aspects of current society, elements of social security and the health care system being the most prominent ones. With respect to the health care system, expenditures are currently about 10% of the Dutch gross domestic product which amounts to about 60 billion guilders per year. Half of these expenditures are on medical services and care services and the largest part of it is consumed by the elderly. In face of this and the trend towards aging, issues concerning the reform of the Dutch health care system are ranked high on the political agenda.

Health status of the elderly has been shown to be a crucial determinant of care services utilization [see 4, 12, 26, 27]. Furthermore, in the Netherlands, providers of care services (either formal or informal) use an individual's health condition as a screening device. Sensible health care politics therefore require understanding of the health status of older persons and of its dynamics preceding death. This paper focuses on these issues. In particular, we give a taxonomy of the health condition of a representative sample of the Dutch elderly population and focus on the forces that drive individuals from good health states to bad health states or to death. Morbidity and mortality rates have often been found to be correlated with socioeconomic status [see 13, 24, 25], gender and living conditions. We will therefore pay attention to the effect of these covariates on morbidity and mortality of the elderly.

Health is not well-defined but it can be safely stated that it is intrinsically a multidimensional concept. There are several types of health disorders [see 3, 6, 7, 15, 19, 23], for instance, some elderly people can experience severe physical limitations while others may have cognitive or emotional impairments. These components have distinct contributions to health and, albeit interrelated, follow different trajectories. They are largely determined by two processes proceeding simultaneously when getting old : a first process which can be referred to as " normal aging" (viz. the usual functional decline in the last phase of the life) and a second one due to age-related diseases. Additionally, while health disorders are more prevalent at advanced ages, the degree of impairment among individuals varies. A large proportion of the Dutch elderly is indeed in good shape whereas others suffer from serious limitations. Multidimensional and graded measures are thus required to describe the health status in the older population.

We use a unique panel survey from the Longitudinal Aging Study Amsterdam

(LASA). The LASA-survey follows a representative sample of Dutch elderly individuals of 55 years and older over time and gathers extensive information on health and factors that may predict changes in health. A large set of indicators is required to cover all dimensions of health. Elements of this set can be seen as indicators of some aspect of health but, used on their own, they are an incomplete predictor of it. It will be difficult to handle all these instruments in any statistical analysis and methods are thus required that summarize this set of measures into a limited number of meaningful scores that can be used in subsequent analyses.

Of the several methods of data reduction that can be used, the Grade of Membership (GoM) method of Woodbury and Manton (1982) [see 18] is the most appropriate method for the problem at hand. GoM is a non-parametric method that identifies simultaneously latent multidimensional profiles of health and the degrees to which each respondent in the sample fits to these profiles. The profiles are meaningful, typical pure health types describing the different dimensions of health. The degrees to which an individual belongs to these types (i.e. grades of membership) are represented by weights that sum up to unity for each individual over all different profiles. So the individual grades of membership are distributed on the interval [0,1] and allow an individual to be partially involved in each dimension of the underlying health concept. These grades can be used in subsequent analyses of health and mortality.

The remainder of this paper is divided into five sections. Section 2 describes the data. The Grade of Membership method and its relation to other, similar techniques are presented in section 3. In section 4 we apply the GoM method on 21 indicators of health observed in two waves of the LASA sample of the Dutch elderly and find a characterization of six pure underlying health types. Subsequent analyses based on the GoM individual membership parameters derived in section 4 are presented in section 5. This section is divided into two subsections. A probability model for mortality will be specified and estimated in the first subsection. We find significant effects of specific health disorders on mortality .and consequently, it is of interest to see which demographic and socioeconomic factors determine changes in these health disorders. Cross-sectional and panel data analyses of health state occupancy are performed in the second subsection. Section 6 summarizes and concludes.

2 Description data set, selection and description of the variables

2.1 Sample

The longitudinal Aging Study Amsterdam (LASA) [see 6, 7] was initiated in 1991 by the Dutch Ministry of Health, Welfare and Sports, Department of Policies for the Aging. The study is intended to lead to policy relevant information on the aging population in the Netherlands and follows a representative sample of non-institutionalized and institutionalized adults older than 55 years over an extended period of time. Currently two waves are available (the 1992/1993 wave and the 1995/1996 wave). Data have been gathered on four components of functioning, physical, emotional, cognitive and social. Each component is assessed by a broad set of objective and subjective instruments, including objective clinical assessments. Information on characteristics that are expected to predict changes in one or more of the components of functioning is also collected. A total of 3107 respondents participated in the first wave of the survey. Of this first wave we select all individuals of 65 years and older. 2141 respondents remain, 1659 of whom also participated in the second wave. Respondents were submitted either to a complete or to a short interview according to their ability to sustain a lengthy interview. Of interest for our analyses is that the bulk of the sample attrition is caused by mortality of respondents in the time intervening between the two waves. 79% of the 482 non respondents to the second wave died in the three years time period.

With respect to the concept of health one may distinguish a physical, a cognitive and an emotional component and the health variables used in the analyses should cover all these dimensions. Below we describe these in more details.

2.2 Instruments for the measurement of health

Data on the functional status of respondents as well as data on their pathological conditions are used. *Physical functioning* is measured by two instruments : three self-report items pertaining to mobility activities in daily life and parts of a performance test of physical ability. The first test assesses the ability of respondents to walk up and down a 15-steps staircase without stopping, to use private or public transport and to cut one's own toenails. We use a total score resulting from this test. The score takes on value 0 when all three activities are performed without difficulties, one, two or three when respectively one, two or three activities are performed with difficulty. However, given the self-reported nature of the test, we are likely to measure the respondent's perception of his ability to perform these activities. Therefore an objective measure of physical ability, namely a performance test, is included to the study. Performance is measured by the time in seconds needed to put on and take off a cardigan. Respondents who could not perform the test are given a score 2. Those with a time score below and above the median are given a score 0 and 1 respectively.

The Mini Mental State Examination (MMSE) is a widely used method for assessing *cognitive status*. It provides a total score that places the individual on a scale of cognitive function. The more cognitive disorders there are, the lower the score is. A cut-off of 23/24 is usually used to indicate cognitive impairment.

Depression is one of the most common disorders of old age and is associated with declines in quality of life and functioning. The Center for Epidemiologic Studies Depression Scale (CES-D) is used to measure *emotional functioning* of the elderly. The total score ranges from 0 to 60 and subjects with a score higher than 16 display clinically relevant symptoms of depression.

Perception is measured by two self-report items on difficulties with seeing and hearing.

The presence of *chronic diseases* is assessed by asking the participants whether they have or have had any of the following diseases : chronic obstructive pulmonary diseases (COPD), heart diseases, atherosclerosis, stroke, diabetes, arthritis, cancer and other chronic diseases. Disease severity should also be taken into account to obtain accurate information about the relation between chronic diseases and disability. Undergoing "-a continuous medical treatment" is a way to assess the severity of a disease. " Continuous medical treatment" was considered present when the patient reported to use prescribed drugs and/or to have regular contacts with a physician for the chronic disease involved [see 15].

We report summary measures of these instruments in table Al of Appendix A.

The set of 21 variables described above captures virtually all aspects of the complex health concept. Unfortunately dealing with these 21 variables in a multivariate analysis is impracticable and it is unclear how to interpret the various possible outcomes of these variables. Therefore methods are needed that comprise the health information in interpretable measures that could be used in additional analyses. The Grade of Membership technique, introduced by Woodbury & Manton (1982) [see, for instance, 18, 19], is a method specifically developed for these purposes. It has been proved to be successful in the gerontological and epidemiological literature. The next section describes this method and relates it to other, similar, techniques such as Principal Component Analysis and Factor Analysis.

3 The Grade of Membership method

Suppose we have a set of observed variables, such as the outcomes of the various tests described above, representing the total information of some underlying health concept that needs to be characterized. Each element of this set of indicators is in itself insufficient to completely characterize this underlying health concept. The difficulty lies in interpreting this multidimensional set of test outcomes and in their use in subsequent analyses. Therefore methods are required that summarize the available information in a limited set of indicators (or an index) that can be interpreted directly and used for other purposes. Methods based on Principle Component Analysis, Factor Analysis or Latent Class Analysis may be practical and suitable for some situations, but are less suitable for our purposes. We will return to this below. We require a method that deals with the multidimensionality of the data, and that recognizes that health of the elderly is a concept of graded participation into several aspects of health. Different types of health disorders exist and partial involvement in these types is possible and should therefore be allowed for. The GoM method, to be described below, is the most suitable candidate to accommodate for this.

GoM identifies simultaneously latent multidimensional profiles and the degrees to which the respondent's features fit these profiles. GoM performs a reparametrization of the variables space : it identifies new dimensions, or latent profiles, so that the attributes of all respondents can be represented as a convex combination of these profiles. The characteristics of these profiles, or, in GoM terminology "pure types", are those of "ideal" respondents manifesting only symptoms associated with a single dimension of the underlying concept. These pure types are not necessarily represented in the sample set. The degrees of similarity between pure types and respondents, viz. grades of membership, are described by weights constrained to fall in the interval [0, 1] and that sum to one over all profiles.

More formally, assume that we have a set of J categorical variables X_j over I individuals. L_j refers to the number of discrete categories l for variable j. Each variable x_{ij} is recoded into a set of L_j binary variables denoted by y_{ijl} where $y_{ijl} = \mathbf{1}$ when the ith respondent has the lth response to the jth variable and $y_{ijl} = 0$ otherwise.

Assume there are K underlying pure health dimensions. The first type of coefficients, denoted by λ_{kjl}, relates the K latent profiles to the J observed variables. λ_{kjl} is defined as the probability that a pure type k has score l for indicator j. A second type of coefficients g_{ik} relates the individual cases to the latent profiles. The individual parameter g_{ik} is the degree of proximity of individual i's characteristics to pure type k's characteristics. The product $g_{ik}\lambda_{kjl}$ is thus the product of the involvement of an individual in a specific profile k and the probability that an individual who belongs entirely to k would have score l for indicator j. The probability that a specific individual responds l to question j, $\text{prob}(y_{ijl}) = 1$, follows then simply as the sum of this product over all possible types of health in which the individual could be involved with. So:

$$\text{prob}(y_{ijl} = 1) = \sum_k g_{ik}\lambda_{kjl}$$

Assuming, conditional on g_{ik} and λ_{kjl}, independence of individual observations, the likelihood function for the model is simply given by :

$$L = \prod_i \prod_j \prod_l (\sum_k g_{ik}\lambda_{kjl})^{y_{ijl}} \tag{1}$$

This likelihood needs to be optimized with respect to the parameters of interest (g_{ik} and λ_{kjl}), subject to the restrictions:

$$\begin{aligned} 0 \leq g_{ik} \leq 1 & \qquad \forall\, i, k \\ \sum_k g_{ik} = 1 & \\ 0 \leq \lambda_{kjl} \leq 1 & \qquad \forall\, k, j, l \\ \sum \lambda_{kjl} = 1 & \end{aligned} \tag{2}$$

Note that the GoM method is basically a non-parametric classification method as no distributional assumption on the form of the distribution of g and λ is required. It may be clear that quite a number of parameters needs to be estimated ((K $-$ 1) times the number of individuals plus $\sum_{j=1}^{J} K * J * (L_j - 1))$ and that maximization of the multinomial likelihood function may be very computer intensive. Briefly we can mention that estimation of GoM parameters rests on the specific structure of the first-order conditions of the likelihood function and that parameters are estimated in an iterative way. More details on the estimation procedure are available on request by the authors. The order of the model (i.e. the number of pure types K) is determined by carrying out a likelihood ratio test on the change in explanatory power following an increase in the number of pure types. The number of grades of freedom equals $(I - 1) + \sum_j L_j$. We refer to Manton & Woodbury (1992) [see 21] for more details.

There are some similarities with other data reduction methods. However, in the GoM method, all parameters are *simultaneously identified* whereas individual parameters in Factor Analysis and Principal Components methods are usually derived after (and thus conditional on) identification of new underlying variables summarizing the information comprised in the data set.

Moreover the GoM method is a non-parametric method and identification of all parameters does not rely on any distributional assumptions. Estimation of factor scores in Factor Analysis rests, for instance, on distributional assumptions concerning the factor loadings.

Additionally, in contrast to Factor Analysis and Principal Components methods, GoM is a classification methodology since respondents on basis of their grades of membership (constrained to fall between **0** and **1)** are allocated to discrete and meaningful groups. Unlike usual classification methodologies (such as for instance cluster analyses), GoM does not create groups of similar entities but takes into account individual heterogeneity. Grades of membership are continuously distributed on the interval [0, 1][1] and allow for partial involvement in one dimension of the underlying concept. GoM is a superior alternative to usual classification methodologies for the problem at hand since decline in health status is also a continuous process : very few old people can be considered as completely disabled but the majority of them retains some type and degree of functioning.

4 GoM analysis on-waves I and II : a typology of health status of the Dutch elderly

The Grade of Membership method is applied on a representative sample of elderly individuals aged 65 and over obtained from the LASA data set and the likelihood function (1) is maximised subject to the set of constraints (2).

4.1 First wave

Table 1 reports the profile probabilities λ_{kjl} for each group k (λ_{kjl} is defined as the probability that a pure type k has score l for indicator j). The individual parameters g_{ik} are the grades of membership and are jointly estimated with parameters λ_{kjl}. The GoM analysis indicates that the latent health concept of the elderly could be well described by six pure types. We tested the order of the model using likelihood ratio tests. The number of degrees of freedom equals 2185. The increase in likelihood value was significant after adding a sixth dimension to the model ($\Delta L_{K=5,K=6} = \mathbf{1849}$) and not significant when adding a seventh dimension ($\Delta L_{K=6,K=7} = 145$ and $L_7 = -3805$, $L_6 = -3950$). This is sufficient to justify the use of six dimensions in the following.

The second column of Table **1** reports sample proportions of various health disorders. The characteristics of the different health dimensions (pure types) are determined by examination of the k profile probabilities and their comparison to sample proportions. The k population frequencies, reported in Table **1,** are given by :

$$f_k = \sum_i g_{ik}$$

[1]Note that there can only be as many GoM scores as there are combinations of answers to the health questions. As a matter of fact there are $2^{\sum_j L_j}$ combinations possible so that in practice the grades of membership can safely be considered as continuous on the interval [0,1].

and roughly indicate the prevalence in the total population of the disorders associated to each pure type.

<insert table 1 about here>

4.2 Typology of the elderly's health status based on results for wave I

- *COPD and cancer patients :* The *first* dimension is characterized by the prevalence of two life threatening diseases : Chronic Obstructive Pulmonary Diseases and cancer. The probability of having a continuous medical treatment, which accounts for the severity of the disease, equals 50 % when having COPD and 35 % when having cancer. Pure type I is also characterized by physical limitations (the score of the self-report test on physical ability is considerably higher than the sample mean) and is more likely to be depressed than others.

- *Other chronic diseases patients :* The health status of the *second* pure type is very good. However profile II is characterized by the presence of "other chronic diseases" : these are mainly diseases which are not specific to the elderly and generally not too serious. Examples of these are hypertension, back troubles or eye diseases.

- *Cognitive impaired :* The *third* pure type is physically healthy but has poor cognitive function (low MMSE test score).

- *Arthritis patients :* The *fourth* pure type is characterized by the prevalence of serious arthritis (almost all respondents follow a continuous medical treatment). Moreover the probability of having an other chronic disease in addition to arthritis is very high. As expected pure type IV is physically impaired (both tests on physical ability account for severe physical limitations). Not surprisingly there is a strong association between being depressed (high CES-D scores) and having serious physical impairments. Finally the cognitive function of pure type IV is also relatively poor (low MMSE test score).

- *Cardiovascular patients :* Difficulties with cardiovascular diseases (heart diseases, atherosclerosis, stroke and diabetes) characterize the *fifth* profile. The vision is much poorer than the mean, likely because of the presence of diabetes and stroke.

- *Healthy elderly people :* Finally the *sixth* dimension is the healthy one.

4.3 Second wave

We next applied the GoM method on **21,** identical, indicators of the second wave (respondents aged **65-85,** n=1736). Results are reported in table B1 of Appendix B. Again we carried out some likelihood ratio tests to justify the order of the model. The increase in likelihood value was significant when adding a sixth dimension and not significant when adding a seventh dimension (the number of degrees of freedom equals 1781, $\Delta L_{K=5\ K=6} = 1455$ and $\Delta L_{K=6,K=7} = \mathbf{117}$**).** The latent health concept is then described by six pure types, similar to those found in the first wave. Consequently there is strong evidence that the typology obtained for wave I covers all aspects of health status (The sample in wave I is representative of the Dutch elderly population [see 6]).

There are several ways to carry out a dynamic analysis. The most common procedure assumes absence of cohort and period effects in small intervals of time, i.e. the content of the dimensions does not change over time. Results of the separate analysis of wave II indicated that this is a valid assumption (see table Bl). Consequently, the profile probabilities of wave I λ^{I}_{kjl} can be assumed to be equal to those of wave II λ^{II}_{kjl}. We next maximise the likelihood function (1) subject to (2) and the additional constraint $\lambda^{I}_{kjl} = \lambda^{II}_{kjl}$. This gives estimates of grades of membership g^2_{ik} for wave II.

As a result of the analyses we comprised the 21 health indicators into 6 underlying variables. For each individual in the sample we now have a characterization of health in both waves by means of grades of membership g^1_{ik} for wave I and g^2_{ik} for wave II. These are used in analyses of mortality and health in the next section.

Before we do that, we first briefly examine the performance of the model to describe the data carrying out an informal diagnostic. A straightforward check is to see whether the model predictions fit the observed data. Estimates of g^t_{ik} and λ_{kjl} could be used to generate $\mathrm{prob}(y_{ijl} = 1)$. A threshold of 0.5 is used to generate predicted $\widehat{y_{ijl}}$ and these are compared with the actual observations y_{ijl}. Table 2 reports the results for wave I and wave II.

<insert table 2 about here>

This simple measure of goodness of fit indicates that 86.6 % in wave I and 86 % in wave II of the observations are predicted correctly. This suggests that the Grade of Membership model fits the observed data quite satisfactorily.

5 Utilization of GoM individual parameters in analyses of mortality and health states occupancy

Before dealing with GoM individual parameters, it is worth recalling that grades of membership g_{ik} indicate the degree of similarity between characteristics of pure type k and respondent i. It has to be noted that GoM is inclined to give weights $g_{ik_1} = 0.5$ and $g_{ik_2} = 0.5$ to individuals having all characteristics of both pure types k_1 and k_2. However, if the combination of these characteristics occurs repeatedly in the sample, it can be expected that GoM will identify this combination as a new dimension of the underlying concept. Consequently, we believe that interpretation remains straightforward : high values of g_{ik} are associated with a high degree of similarity of respondent i to pure type k. We propose to use these individual parameters in two related analyses : analysis of mortality and analyses of the demographic and socioeconomic factors determining membership of specific health dimensions.

5.1 Analysis of mortality

For the analysis of mortality rates, a logit analysis is performed with as dependent variable the binary variable "deceased" (0), "not deceased" (1). 21 % of the respondents has died between the two waves. We relate mortality to health states occupied in wave I and to a range of demographic and socioeconomic characteristics. Details on these variables are given in table A2 of Appendix A. Results of the logit analysis on mortality are reported in Table 3.

<insert table 3 about here>

We find strong effects of membership of profiles I, IV and V as well as of age, gender and income on mortality. The health pure types I (COPD/cancer) and V (cardiovascular diseases) are both life threatening and occupancy of these health dimensions increases the probability of dying. The disease associated with profile IV (arthritis) is not life threatening but is often associated with poor physical, emotional and cognitive health and respondents with a high grade of membership in profile IV face increased mortality risks. Additionally we find a significant positive effect of age after we control for health. It is also well-known that females have a higher life expectancy than males and this is reflected in higher mortality rates for males. Higher incomes are associated with lower mortality rates. It is conceivable that individuals with higher socioeconomic status have experienced better life style and/or better living conditions than their counterparts, and are, in consequence, inherently healthier. Therefore, they may get a life threatening disease later in life and, when having a disease, they may survive longer.

It is important to note that the health dimensions I, IV and V are the most dominant factors explaining mortality. It is therefore crucial to see which factors determine membership of these profiles.

5.2 Analyses of heath states occupancy

In the next two subsections we are interested in relating membership of specific health dimensions to observed demographic and socioeconomic characteristics.

5.2.1 Cross-sectional analyses

As a first step the following linear model is estimated on wave I.

$$g_{ik}^1 = x_i' \beta_k^1 + u_{ik}^1 \tag{3}$$

for $i = 1, \ldots, I$, $k = 1, \ldots, 5$.
Simple OLS estimates based on the grades of membership derived from wave I g_{ik}^1, for $k = 1, \ldots, 6$, are reported in table 4.

<insert table 4 about here>

We find significant effects of most demographic and socioeconomic variables. With respect to age and gender, we find that aging is associated with increased probability of poor health. Remarkably females, who are known to have a higher life expectancy than males (see, for instance, table 3), are observed to experience more often health disorders than males. This may be explained by two additional results of our analyses : males have higher risks of incidence of cardiovascular disorders (life threatening diseases) and females suffer more often than males from arthritis and other chronic diseases (non life threatening diseases). Additionally living in urban areas seems to be associated with aggravation of respiratory diseases and cancer. There is also some evidence that living in an one-person household is associated with emergence or aggravation of health disorders. Finally positive effects of education and income on health are found : high levels of education and/or income are associated with better health conditions. In particular, absence of arthritis and cardiovascular diseases (pure types IV and V) seems to be related with high incomes and education levels. Also lower risks of suffering from cognitive impairments are found under the more highly educated individuals.

5.2.2 Panel data analyses with correction for attrition due to mortality

The LASA data set consists of two waves and both waves were used to construct individual GoM parameters. The longitudinal nature of the data set could be exploited explicitly. Our data concern a survey of the older population from which a significant fraction of the initially selected respondents is observed to die between the two waves. A common practice if attrition occurs in a sample is to discard individuals for whom data is missing. Consequently only information on individuals who did not drop out of the sample is used. It is then implicitly assumed that mortality has no influence on the parameters of interest. This assumption may be untenable in the situation at hand since it can for instance be argued that individuals with higher than average grades in health dimensions ***I, IV*** and ***V*** at wave I may be expected to have a higher mortality rate as they have a worse health condition than others. These will be underrepresented in the balanced sample. Reciprocally high grades in profiles ***II*** and ***VI*** might protect against death. We firstly tested for non random selection carrying out a simple diagnostic. Cross-sectional regressions are performed on wave I in a similar way as in 5.2.1 with as additional dependent variable the binary index " deceased" (1), "not deceased" (0) in wave II. Significance of the binary indicator signals non random attrition due to mortality. See [8] for details. Results regarding this extra regressor only are reported in table 5.

<insert table 5 about here>

The variable " deceased" , "not deceased" is significant for all health dimensions and this argues for the presence of non random attrition in the sample. Moreover the parameters confirm that, in the balanced sample, individuals with high grades of membership in profiles II and VI are overrepresented and individuals with high grades of membership in profiles I, IV and V are underrepresented. Consequently inferences based on a balanced sample will lead to biased estimates of the parameters of interest.

In line with this a model could be written in which mortality and health states occupancy are determined jointly. Consequently we estimated a sample selection model along the lines of Hausman and Wise [see 10] where death and grades of membership are assumed to depend on linear functions of a range of demographic and socioeconomic characteristics. This model can be written as :

$$\begin{aligned}
A4^* &= x^1\delta + \omega \qquad (4)\\
g_1^1 &= x^1\beta_1 + l_1\alpha + u_1^1\\
g_1^2 &= x^2\beta_1 + l_1\alpha + u_1^2\\
g_2^1 &= x^1\beta_2 + l_2\alpha + u_2^1\\
g_2^2 &= x^2\beta_2 + l_2\alpha + u_2^2\\
g_3^1 &= x^1\beta_3 + l_3\alpha + u_3^1\\
g_3^2 &= x^2\beta_3 + l_3\alpha + u_3^2\\
g_4^1 &= x^1\beta_4 + l_4\alpha + u_4^1\\
g_4^2 &= x^2\beta_4 + l_4\alpha + u_4^2\\
g_5^1 &= x^1\beta_5 + l_5\alpha + u_5^1\\
g_5^2 &= x^2\beta_5 + l_5\alpha + u_5^2
\end{aligned}$$

where M^* is a latent continuous variable measuring the inclination of staying alive, i.e of staying in the panel. Individuals drop out of the panel (die) if the latent index $M^* \leq \mathbf{0}$. The sixth health dimension is excluded from the model to avoid perfect correlation between the dependent variables $(\sum_k g_{ik} = 1)$. u_k^1, u_k^2 and ω are assumed to be normally distributed with zero means and to be mutually independent. It is conceivable that time-invariant unobservables play an important role in analyzing health since some people may be intrinsically healthier than others (for instance due to genetic factors or different life styles). Therefore in line with this we control for presence of unobservables α.[2] α is assumed to be normally distributed with zero mean and variance σ_α^2, and is independent of u_k^t, for $(t=1,2)$ and $(k=\mathbf{1},\ldots,5)$. α is allowed to be correlated with w. The covariance between α and w, denoted by $\sigma_{\alpha\ \omega}^2$ is crucial for the problem at hand, as $\sigma_{\alpha\ \omega}^2 = 0$ implies that attrition due to mortality and grades of membership are independently distributed and that hence attrition between wave I and II has no influence on the parameters of interest. More details on the estimation procedure are given in Appendix C.

Table 6 reports results of estimation of model (4).

<insert table 6 about here>

In the first part of table 6, the first column reports the results for mortality (coefficients δ), the five following columns reports results for the health variables g_k (coefficients β_k). Parameters for the sixth health dimension can easily be derived using β_k, for $k = \mathbf{1},\ldots,5$ as $\sum_k g_{ik} = \mathbf{1}$. The second part gives some results concerning the variance-covariance matrix of the residuals $(l_k$, u_k^t, for $t=1,2$ and $k=\mathbf{1},\ldots,5$ and a).
To begin with the last part of the table, we can see that $\sigma_{\alpha\ \omega}^2$ is significant, implying that mortality and health are stochastically related and that exclusion of deceased respondents in panel data analyses of g_{ik} will lead to biased estimates. Consequently attrition due to mortality must be taken into account in the estimation of the health model. The variance σ_α^2 of individual effects is highly significant which accounts for the existence of time-persistent unobservables.
With respect to the results for mortality (δ), negative coefficients are associated with higher mortality rates. The results are in line with the analyses of subsection **5.1.** Age, income and gender have important significant effects on mortality. Females have lower mortality rates. High socioeconomic status is associated with lower mortality. Note however that effects of regressors in this model are essentially reduced form effects and are difficult to compare directly.
The parameters concerning health are briefly discussed as most of these are very much in line with the results of subsection 52.1. It is worth pointing out that the

[2]We assumed a so called one-factor error model for the unobservables in the health equations and, as required for the estimation procedure, normalized l_1 to one.

effects of time varying regressors may differ from those in the cross-sectional analyses. The effects of variables of subsection **5.2.1** are identified from interindividual variation whereas the analyses in this subsection use both inter- and intraindividual variations. In **5.2.1,** we measure for instance effects of the living situation (cf "alone") on health whereas, in **5.2.2,** effects of changes in the living situation are observed. These effects differ substantially in both analyses which suggests that changes in living conditions have more impact on the health of the elderly than the living situation itself.

Once again significant effects of age, gender, socioeconomic status, urbanisation and of living alone on health are found. In particular, a high socioeconomic status has a substantial positive effect on the health condition. It is worth noting that individuals with high incomes face higher risks of emergence of aggravation of cancer or COPD diseases.

To conclude it has to be noted that, irrespective of the fact that individual effects and attrition due to mortality should be taken into account in analyses on health, estimates of parameters of models (3) and (4) are reasonably comparable.

6 Summary and conclusion

The paper shows that the Grade of Membership method is the most suitable method to summarize health, a multidimensional and dynamic concept, into a limited number of indices that can be used in subsequent analyses on health and mortality.

In the first part of this paper we focus on methods that comprise multidimensional health measures into a limited number of indices. The Grade of Membership approach introduced by Manton and Woodbury in 1982 is specifically designed to characterize the complex concept of health. The method simultaneously identifies all dimensions of the concept of interest and the degrees to which an individual belongs to each of these profiles. We apply the method on a set of 21 indicators from the Longitudinal Aging Study Amsterdam data set. Six dimensions have been identified which can be briefly characterised as patients suffering from COPD and cancer (type I), from other chronic diseases (type II), from arthritis (type III), from cardiovascular diseases (type IV), cognitive impaired elderly individuals (type V) and healthy old people (type VI).

These findings are however not easy to compare with other Grades of Membership analyses of the health status of elderly populations. The typology obtained strongly relies on the choices of measurement instruments and not all studies use the same instruments nor do these instruments capture similar aspects of the health concept. However it can for example be mentioned that the GoM analysis of disability of the U.S. elderly (Manton, Stallard and Corder [see 23]) indicates that the health status of the elderly can be well described by six dimensions characterized as healthy and unimpaired (type I), some physical limitations (type II), some physical limitations due to cardiopulmonary problems (type III), cognitive limitations often due to dementia and stroke (type IV), moderate physical impairments often due to musculoskeletal problems (type V) and complexly impaired (type VI). These dimensions are to a large extent comparable with our typology derived above.

The remainder of the paper uses the simultaneously derived set of grades of membership to analyse mortality and transitions in health states. As far as mortality is concerned, besides strong expected effects of aging, major effects of specific health states, namely COPD, cancer and cardiovascular diseases, and of socioeconomic status (measured by education and income) have been found. We then estimated a range of models to analyse health state occupancy and find that it is important to analyse mortality and health jointly in order to avoid biases in the parameter estimates of a model for health dynamics. In addition to effects of aging on health, we find four major additional effects. Firstly females, although having a longer life expectancy, experience health disorders more often than males (type VI). This may be explained by the fact that a higher incidence of cardiovascular disorders (life threatening diseases) is observed for males (type V) whereas females suffer more often than males from non life threatening diseases, i.e. arthritis and other chronic diseases (type VI). Finally strong positive effects of socioeconomic status on health are shown (type VI). Explanation of this result needs further research.

These findings are in line with previous work in the gerontological literature. For instance, [7, 25] also found the prevalence of chronic diseases to increase with age, higher incidence of cardiac diseases among men and higher incidence of arthritis among women. Age, levels of education and female gender have also been observed to be significant determinants of cognitive decline. The positive effect of socioeconomic status on mortality has also been found in for instance [13, **14,** 24].

Results obtained using Grades of membership methods may contribute in various ways to the debate concerning policy problems posed by aging populations. On one hand the typology of health status derived using GoM methods can be used to characterize heterogeneous elderly populations and to follow their evolution given foreseeable demographic trends. On an other hand, the individual grades of membership permit a micro analysis of the determinants of major health disorders as well as of their development in the future.

Then studies on mortality of the elderly using GoM methods may help providers of care services in determining frail populations, i.e. populations with a high probability of dying and consequently populations with, often, higher needs for care services.

Moreover, most studies on future needs in health care services [see, for instance, 4, 26, 27] rely on present and past utilization. Given that the supply side of the market in the Netherlands is highly regulated, the actual utilisation of care services does not properly describe the true needs of the elderly. Having assessed the health status of Dutch elderly individuals, and its changes over time, using GoM methods, it is then possible to relate it to preferences of the elderly for care facilities and estimate more accurately, given the foreseeable demographic trends, the future potential need for care services.

Finally, for given health status, it will also be possible to assess the possible discrepancies between, on one hand, preferences and needs of the elderly for care facilities and, on the other hand, actual use of care facilities.

Table **1** *: Results of Grade of Membership Analysis on the first wave of the LASA data. Estimates* ***of*** *probabilities* (λ_{kjl}) *describing the six health dimensions identified from* ***21*** *health measures.*

λ_{kjl}		freq.	PURE TYPES: COPD, Cancer patients K=1	Other chronic diseases K=2	Cognit. impaired K-3	Arthritis patients K-4	Cardio vascular diseases K-5	Healthy elderly people K-6
Pop. freq. (%)			11.2	10.5	12.0	15.1	17.7	33.5
Self-report test	0	0.47		1			0.41	1
on physical ability	1	0.22	0.18		1	0.02	0.17	
	2	0.15	0.81			0.37		
	3	0.14				0.61	0.42	
Performance test	0	0.37	0.41	0.44	0.45	0.09	0.15	0.56
(cardigan)	1	0.59	0.59	0.55	0.49	0.78	0.85	0.44
	2	0.04			0.05	0.13		
MMSE	0	0.85	1	1	0.68	0.67	0.79	1
	1	0.15			0.32	0.33	0.21	
CES-D	0	0.82	0.66	0.96	0.86	0.33	0.72	0.80
	1	0.16	0.34	0.04	0.14	0.67	0.28	0.20
Vision	0	0.87	1	1	1	1	0.42	1
	1	0.13					0.58	
Hearing	0	0.87	0.93	0.94	1	0.62	0.85	0.95
	1	0.10	0.07	0.06	-	0.38	0.15	0.05
COPD	N	0.86	0.33	1	1	1	1	1
	Y	0.13	0.67		-			
Med. treat. (COPD)	N	0.90	0.50	1	1	1	1	1
	Y	0.10	0.50	-	-	-		
Heart diseases	N	0.77	1	1	1	1		0.81
	Y	0.23		-	-	-	1	0.19
Med. treat. (Heart d.)	N	0.79	1	1	1	1		1
	Y	0.20	-	-	-	-	1	
Atheroscler.	N	0.88	1	1	1	1	0.39	1
	Y	0.11	-	-	-	-	0.61	-
Med. treat. (Athero.)	N	0.92	1	1	1	1	0.67	1
	Y	0.08	-	-	-	-	0.33	-
Diabetes	N	0.89	1	1	1	1	0.50	1
	Y	0.10	-	-	-	-	0.50	-
Med. treat. (Diab.)	N	0.91	1	1	1	1	0.60	1
	Y	0.09	-	-	-		0.40	

λ_{kjl}		freq.	PURE TYPES					
			K=1	K=2	K=3	K=4	**K=5**	**K=6**
Stroke	N	0.92	1	1	1	1	**0.67**	**1**
	Y	0.07				-	0.33	-
Med. treat. (Stroke)	N	0.95	1	1	1	1	**0.74**	**1**
	Y	0.05				-	0.26	-
Arthritis	N	0.63	0.53	0.72	0.64	-	0.64	1
	Y	0.37	0.47	0.28	0.36	1	**0.36**	-
Med. treat. (Arthritis)	N	0.83	1	1	1	0.06	1	1
	Y	**0.16**				**0.94**	-	-
Cancer	N	0.88	0.38	1	1	1	1	1
	Y	**0.11**	0.62	-	-			
Med. treat. (Cancer)	N	0.93	0.65	1	1	1	1	1
	Y	0.07	0.35	-	-	-	-	-
Other chronic diseases	N	0.65	1		1		1	1
	Y	0.34		1		1		

Table 2 : Goodness of fit of GoM, wave I and wave II

		Predicted (in %)			
		Wave I		*Wave II*	
		P=0	P=1	P=0	P=1
Actual (in %)	P=0	47.0	7.1	47.0	7.8
	P=1	6.3	39.6	6.2	39.0

Table 3 : Estimates of a logit analysis on mortality (deceased=0).

	Logit coefficients	(t-student)
g_1^1	-1.73	(-4.72)
g_2^1	-0.13	(-0.16)
g_3^1	0.05	(0.18)
g_4^1	-0.81	(-2.44)
g_5^1	-2.13	(-6.34)
g_6^1	-0.32	(-0.42)
Age	-0.46	(-6.50)
Education	0.03	(1.46)
Income	0.09	(3.15)
Urbanisation degree	0.10	(0.82)
Female	0.81	(5.18)
Alone	-0.24	(-1.32)

Table 4 : OLS estimation results of wave I (P-parameters, T-values in parentheses).

	PURE TYPES					
	COPD, Cancer patients	Other chronic diseases	Cognit. impaired	Arthritis patients	Cardio vascular diseases	Healthy elderly people
	g_1	g_2	g_3	g_4	g_5	g_6
Constant	0.07	0.13	0.07	-0.04	0.17	0.58
	(2.61)	(3.70)	(2.37)	(-1.33)	(4.77)	(10.21)
Age	0.01	-0.02	0.01	0.03	0.03	-0.06
	(2.45)	(-5.12)	(3.57)	(7.50)	(8.53)	(-9.96)
Education	-0.01	0.005	-0.01	-0.01	-0.01	0.03
	(-2.04)	(1.06)	(-2.05)	(-2.10)	(-2.09)	(4.10)
Income	-0.005	0.01	-0.007	-0.01	-0.01	0.03
	(-0.74)	(1.03)	(-0.99)	(-1.44)	(-2.08)	(2.47)
Urbanis. degree	0.01	-0.002	-0.002	-0.005	-0.001	0.005
	(2.34)	(-0.42)	(-0.35)	(-0.92)	(-0.10)	(0.55)
Female	0.004	0.025	0.01	0.07	-0.03	-0.08
	(0.49)	(2.41)	(1.47)	(7.69)	(-3.43)	(-5.12)
Alone	0.001	0.007	-0.004	0.01	-0.01	-0.01
	(0.09)	(0.59)	(-0.45)	(1.47)	(-0.93)	(-2.52)

Table 5 : OLS estimation results of wave I (&parameters, T-values in parentheses) with additional regressor "deceased", "not deceased" (excerpt).

	PURE TYPES					
	COPD, Cancer patients	Other chronic diseases	Cognit. impaired	Arthritis patients	Cardio vascular diseases	Healthy elderly people
	g_1	g_2	g_3	g_4	g_5	g_6
Deceased in wave II	0.04	-0.04	-0.02	0.02	0.09	-0.10
	(4.42)	(-3.09)	(-1.71)	(2.02)	(7.07)	(-5.15)

Table 6 : Results of a panel data model for g_{ik} and mortality (δ and β parameters, T-values in parentheses).

	Mortality equation (deceased=0)	COPD Cancer patients	Other chronic diseases	Cognit. impaired	Arthritis patients	Cardio vascular diseases
Cst	0.808	0.070	0.168	0.102	-0.050	0.184
	(3.049)	(2.463)	(6.791)	(4.170)	(-1.756)	(6.487)
Age	-0.341	0.006	-0.019	0.007	0.028	0.036
	(-10.03)	(1.823)	(-6.416)	(2.399)	(8.055)	(10.53)
Education	0.098	-0.014	0.009	-0.004	-0.009	-0.008
	(2.619)	(-3.297)	(2.741)	(-1.271)	(-2.381)	(-2.045)
Income	0.153	0.008	-0.005	-0.007	-0.004	-0.019
	(2.428)	(1.615)	(-0.935)	(-1.299)	(-0.692)	(-2.850)
Urb. Deg.	-0.005	0.011	-0.004	-0.005	-0.002	0.002
	(-0.129)	(2.187)	(-1.192)	(-1.340)	(-0.634)	(0.517)
Female	0.392	-0.000	-0.002	0.012	0.084	-0.042
	(5.097)	(-0.016)	(-0.404)	(1.749)	(10.05)	(-5.083)
Alone	0.140	0.019	-0.007	-0.003	0.031	-0.007
	(1.671)	(2.482)	(-0.942)	(-0.429)	(3.522)	(-0.808)
Variance-covariance residuals :						
l_k		1	-0.178	-0.231	-0.061	0.000
			(-7.576)	(-9.949)	(-2.258)	(0.019)
$\sigma^2_{u_k}$		0.008	0.051	0.036	0.044	0.048
		(10.93)	(29.77)	(30.14)	(29.81)	(30.32)
$\sigma^2_{u_k}$		0.014	0.025	0.030	0.052	0.047
		(16.65)	(26.42)	(26.71)	(26.46)	(27.03)
σ^2_{α}			0.026			
			(23.13)			
$\sigma_{\alpha,\omega}$			-0.030			
			(-4.072)			

References

[1] M. van Baal *Figures from health surveys ; health indicators 1981-1995 (in Dutch),* Maandbericht gezondheidsstatistiek,
Centraal Bureau voor de Statistiek, july 1996.

[2] C Chatfield and A. J. Collins *Introduction to multivariate analysis,*
Chapman and Hall Ltd, 1980.

[3] A T. F. Beekman *Depression in later life : Studies in the community,* Dissertation Vrije Universiteit
CopyPrint 2000, Enschede, the Netherlands, 1996.

[4] Centraal Bureau voor de Statistiek *Statistical handbook of the elderly (in Dutch),*
Centraal bureau voor de Statistiek, 1990.

[5] D. J. H. Deeg *Experiences from longitudinal studies of aging,*
Netherlands Institute of Gerontology, 1989.

[6] D. J. H. Deeg, M. Westendorp de Seriere, A. T. F. Beekman and D. M. W. Kriegsman *Autonomy and well-being in the aging population, report from the Longitudinal Aging Study Amsterdam 1992-1993,*
VU University Press, Amsterdam 1994.

[7] D. J. H. Deeg, M. Westendorp de Serière, A. T. F. Beekman and D. M. W. Kriegsman *Autonomy and well-being in the aging population 2, report from the Longitudinal Aging Study Amsterdam 1992-1996,*
VU University Press, Amsterdam 1998.

[8] J. Fitzgerald, P. Gottschalk and R. Moffitt *An analysis of sample attrition in panel data,*
Journal of Human Ressources (Spring 1998), Vol. 33.

[9] R L. Gorsuch *Factor Analysis,*
Lawrence Erlbaum Associates Publishers, 1983.

[10] J. A. Hausman and D. A. Wise *Attrition bias in experimental and panel data : the Gary income maintenance experiment,*
Econometrica (March 1979), Vol. 47.

[11] C. Hsiao *Analysis of Panel Data,*
Cambridge University Press, 1986.

[12] R. Huijsman *A model for care facilities for the elderly (in Dutch),*
Thesis Rijksuniversiteit Limburg, 1990.

[13] M. D. Hurd, D. McFadden and A. Merrill *Healthy, Wealthy, and Wise? Socioeconomic status, morbidity, and mortality among the elderly,*
Rand Corporation, Workingpaper, 1998.

[14] G. A. Kaplan, T. E. Seeman et, al. *Mortality among the elderly in the Alameda County Study : behavioral and demographic risks factors,*
American journal of public health, 1987, Vol. 77.

[15] D. M. W. Kriegsman *Chronic Diseases, Family Features and Physical Functioning in Elderly People,*
Thesis Publishers, 1995.

[16] G. S. Maddala *The econometrics of Panel Data,*
Edward Elgar Publishing Limited, 1993.

[17] G. S. Maddala. *Limited-dependent and qualitative variables in econometrics,*
Cambridge University Press, 1985.

[18] K. G. Manton and M. A. Woodbury *A new procedure for analysis of medical classification,*
Methods of Information in Medicine 1982, Vo1.21.

[19] K. G. Manton *A longitudinal study of functional change and mortality in the United States,*
Journal of Gerontology : Social sciences 1988, Vol. 43.

[20] K. G. Manton and E. Stallard *Cross-sectional estimates of active life expectancy for the U.S. elderly and oldest-old populations,*
Journal of Gerontology : Social sciences 1991, Vol. 46.

[21] K. G. Manton, M. A. Woodbury, E. Stallard and L. S. Corder *The use of Grade of Membership techniques to estimate regression relationships,*
Sociological Methodology 1992.

[22] K. G. Manton, D. Wolf, J. Ondrich, E. Stallard, M. A. Woodbury and L. Corder *A model for simulating life histories of the elderly : Model Design and Implementation Plans,*
Maxwell Center for Demography and Economics -of Aging, August 1995.

[23] K. G. Manton, E. Stallard *The dynamics of Dimensions of age-related disability 1982 to 1994 in the U.S. elderly population,*
Journal of Gerontology : Biological sciences 1998, Vol. 53a.

[24] T. Martelin *Mortality by indicators of socioeconomic status among the Finnish elderly,*
Social Sciences Medicine, 1994, Vol. 38.

[25] S. Robert and J.H. House *SES differentials in health by age and alternative indicators of SES,*
Journal of aging and health, 1996, Vol. 8.

[26] Sociaal en Cultureel Planbureau *A forecasting model for the health care services (in Dutch),*
Centraal Planbureau, 1997.

[27] Stuurgroep Toekomstscenario's Gezondheidszorg *The elderly population in 2005 : Health and Care (in Dutch),* Stuurgroep Toekomstscenario's Gezondheidszorg, 1992.

Appendix A : Description of variables.

Al : Health variables.

	Score	Freq. in % (Wave I)	Freq. in % (Wave 11)
Self-report test	*0*	0.47	0.44
on physical ability	1	0.22	0.26
	2	0.15	0.15
	3	0.14	0.14
Performance test	0	0.37	0.53
(cardigan)	1	0.59	0.44
	2	0.04	0.02
MMSE (cut-off = 24)	0	0.85	0.87
	1	0.15	0.13
CES-D (cut-off = 16)	0	0.82	0.85
	1	0.16	0.15
Vision	0	0.87	0.89
	1	0.13	0.11
Hearing	0	0.87	0.88
	1	0.10	0.12
COPD	N	0.86	0.85
	Y	0.13	0.15
Med. treat. (COPD)	N	0.90	0.85
	Y	0.10	0.15
Heart diseases	N	0.77	0.74
	Y	0.23	0.26
Med. treat. (Heart dis.)	N	0.79	0.74
	Y	0.20	0.26
Atheroscler.	N	0.88	0.86
	Y	0.11	0.14
Med. treat. (Athero.)	N	0.92	0.88
	Y	0.08	0.12
Diabetes	N	0.89	0.91
	Y	0.10	0.09
Med. treat. (Diab.)	N	0.92	0.91
	Y	0.09	0.09
Stroke	N	0.92	0.91
	Y	0.07	0.09
Med. treat. (Stroke)	N	0.95	0.92
	Y	0.05	0.08
Arthritis	N	0.63	0.51
	Y	0.37	0.49
Med. treat. (Arthritis)	N	0.63	0.54
	Y	0.16	0.46
Cancer	N	0.88	0.86
	Y	0.11	0.14
Med. treat. (Cancer)	N	0.93	0.88
	Y	0.07	0.12
Other chronic dis.	N	0.65	0.75
	Y	0.34	0.25

: ***Demographic and socioeconomic variables.***

	Cut-offs	Scores	Freq. (in %) (Wave I)	Freq. (in %) (Wave II)
Age	1923	1	1.26	16.03
(year of birth)	1918-1922	2	22.7.4	22.89
	1913-1917	3	22.60	20.18
	1908-1912	4	26.85	21.62
	1903-1907	5	27.09	15.60
Education	Elementary not completed	1	11.77	10.03
	Elementary completed	2	38.25	34.25
	Lower vocational education	3	18.03	18.97
	More than lower vocat. educ.	4	31.66	36.62
Income in guilders per month	O-1750	1	31.71	23.99
	1751-2500	2	21.90	29.01
	2501-4000	3	17.56	22.95
	>4000	4	9.20	13.03
	(Missing values)		14.61	11.01
Urbanisation level	Low	1	29.61	30.39
	Medium	2	30.35	30.50
	High	3	40.02	39.10
Sex	Male	1	48.52	46.65
	Female	2	51.47	53.34
Alone	Not living alone	1	36.06	34.47
	Living alone	2	63.94	65.52

Appendix B : Results of a Grade of Membership analysis (second wave).

Table Bl : Results of GoM Analysis on the second wave of the LASA data. Estimates of the probabilities (λ_{kjl}) describing the six health dimensions identified from the health measures. (λ_{kjl} unrestricted)

λ_{kjl}		freq.	PURE TYPES COPD Cancer patients K=l	Other chronic diseases K=2	Cognit. impaired K - 3	Arthritis patients K-4	Cardio vascular diseases K-5	Healthy elderly people K-6
Pop. freq (%)			13.4	72 .	99 .	21.1	17.6	30.7
Self-report test	0	0.44	0.28	0.72			0.13	1
on physical ability	1	0.26	0.09	0.10	1	0.18	0.21	
	2	0.15	0.63	0.17	-	0.26	0.25	
	3	0.14		-	-	0.56	0.41	
Performance test	0	0.53	0.60	0.74	0.37	0.29		0.64
(cardigan)	1	0.44	0.26	0.24	0.61	0.54	0.98	0.36
	2	0.02	0.14	-	0.02	0.17	0.02	
MMSE	0	0.87	0.94	1	0.61	0.83	0.67	1
	1	0.13	0.06	-	0.39	0.17	0.33	
CES-D	0	0.85	0.90	1	0.91	0.60	0.71	1
	1	0.15	0.10	-	0.09	0.40	0.29	
Vision	0	0.89	1	1	1	1	0.38	1
	1	0.11		-	-		0.62	
Hearing	0	0.88	1	0.79	1	0.56	0.86	0.79
	1	0.12		0.21	-	0.44	0.14	0.21
COPD	N	0.85	0.07	1	1	1	1	1
	Y	0.15	0.93	-	-			
Med. treat (COPD)	N	0.85	0.26	1	1	1	1	1
	Y	0.15	0.74	-	-			
Heart	N	0.74	0.53	1	1	1		0.93
	Y	0.26	0.47	-	-		1	0.07
Med. treat (Heart d.)	N	0.74	1	1	1	0.96		1
	Y	0.26		-	-	0.04	1	
At hero	N	0.86	1	1	1	1	0.08	1
	Y	0.14	-	-	-		0.92	
Med. treat (Athero.)	N	0.88	1	1	1	1	0.40	1
	Y	0.12	-	-	-	-	0.60	

λ_{kjl}			PURE TYPES					
		freq.	K=1	K=2	K-3	K - 4	K - 5	K - 6
Diabetes	N	0.91	1	1	1	1	0.44	1
	Y	0.09				-	0.56	-
Med. treat (Diab.)	N	0.91	1	1	1	1	0.38	1
	Y	0.09				-	0.62	-
Stroke	N	0.91	1	1	1	1	0.67	1
	Y	0.09				-	0.33	-
Med. treat (Stroke)	N	0.92	1	1	1	1	0.66	1
	Y	0.08				-	0.34	-
Arthritis	N	0.51	1	0.04	1		1	1
	Y	0.49	-	0.96	-	1		
Med. treat (Arthrit.)	N	0.54	1	0.71	1	-	0.09	1
	Y	0.46	-	0.29	-	1	0.91	-
Cancer	N	0.86	0.42	1	1	1	1	1
	Y	0.14	0.58	-	-			
Med. treat (Cancer)	N	0.88	0.44	1	1	1	1	1
	Y	0.12	0.56	-	-			
Other chronic dis.	N	0.75	1		1	0.24	1	1
	Y	0.25		1		0.76	-	-

Appendix C : Derivation of the likelihood function of model (4) explaining health states occupancy with correction of mortality.

The contribution to the likelihood of the ith observation is given by :

$$L(i) = \Leftrightarrow \begin{cases} f(g_{i1}^1, g_{i1}^2, g_{i2}^1, g_{i2}^2, g_{i3}^1, g_{i3}^2, g_{i4}^1, g_{i4}^2, g_{i5}^1, g_{i5}^2, M_i) & \text{if } M_i = 1 \\ f(g_{i1}^1, g_{i2}^1, g_{i3}^1, g_{i4}^1, g_{i5}^1, M_i) & \text{if } M_i = 0 \end{cases}$$

so the likelihood for the whole sample is

$$\begin{aligned} L(i) &= \prod_{M_i=1} f(g_{i1}^1, g_{i1}^2, g_{i2}^1, g_{i2}^2, g_{i3}^1, g_{i3}^2, g_{i4}^1, g_{i4}^2, g_{i5}^1, g_{i5}^2, M_i) \prod_{M_i=0} f(g_{i1}^1, g_{i2}^1, g_{i3}^1, g_{i4}^1, g_{i5}^1, M_i) \qquad (5) \\ &= \prod_{M_i=1} f(M_i/g_{i1}^1, g_{i1}^2, g_{i2}^1, g_{i2}^2, g_{i3}^1, g_{i3}^2, g_{i4}^1, g_{i4}^2, g_{i5}^1, g_{i5}^2) f(g_{i1}^1, g_{i1}^2, g_{i2}^1, g_{i2}^2, g_{i3}^1, g_{i3}^2, g_{i4}^1, g_{i4}^2, g_{i5}^1, g_{i5}^2) \\ &\quad \prod_{M_i=0} f(M_i/g_{i1}^1, g_{i2}^1, g_{i3}^1, g_{i4}^1, g_{i5}^1) f(g_{ik}^1, g_{i2}^1, g_{i3}^1, g_{i4}^1, g_{i5}^1) \end{aligned}$$

The univariate conditional densities in likelihood (5) are derived using standard results from statistics, using the covariance structure of the residuals.